ARO 9409.3 M; 12939.14m
$13.34.12$-m

LEVEL Ⅲ

⑫

BAYESIAN METHODS, FORECASTING AND CONTROL IN
STATISTICS AND OPERATIONS ANALYSIS

A FINAL REPORT

BY

G.E.P. BOX AND G.C. TIAO

DECEMBER 5, 1981

DTIC
SELECTED
JAN 4 1982
H

TO

U.S. ARMY RESEARCH OFFICE

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON

8112 31101

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A109199	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Bayesian Methods, Forecasting and Control in Statistics and Operations Analysis		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G.C. Tiao G.E.P. Box		8. CONTRACT OR GRANT NUMBER(s) DA-ARO-D-31-124-72-G162 DAHC04-76-G-0010 DAAG29-76-G-0304 DAAG29-78-G-0166
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Wisconsin-Madison 1210 W. Dayton St. Madison, WI 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE December 1, 1981
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

time series, ARMA models, seasonal adjustment, temporal and contemporal aggregation, intervention analysis, outliers, statistical inference and criticisms, robustness.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this final report, a summary of main results is given for research in the following areas:

(a) development of techniques for the analysis and decomposition of seasonal time series
(b) modeling and analysis of univariate and multiple time series
(c) robustness in statistical analysis
(d) a unified theory of statistical inference and criticism.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Statement of Problems Studied
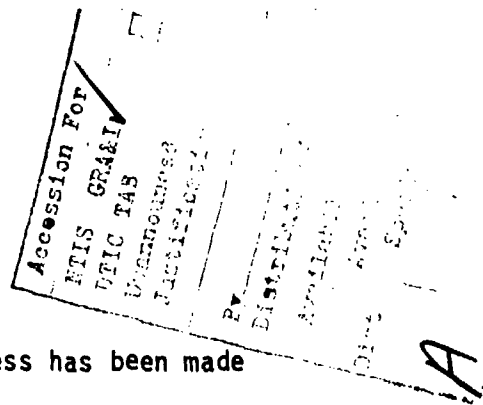
During the last nine years, considerable progress has been made by the principal investigators and their associates in the development of Bayesian methods and time series techniques under the sponsorship of this project. In this final report, we shall provide a brief summary of the main results of the following principal areas of research: (i) analysis and decomposition of seasonal time series, (ii) effect of temporal and contemporal aggregation of time series, (iii) intervention techniques in time series, (iv) modelling and analysis of multiple time series, (v) asymptotic properties of least squares estimates in time series and extended autocorrelation function, (vi) pooling of time series, (vii) effect of outlier in time series, (viii) science and statistics, (ix) robustness in statistical analysis, (x) unified theory of statistical criticism and inference.

Throughout this report, we shall frequently refer to the following univariate and multiple time series models. Let $\{Z_t\}$ be a series of observations taken at equal spaced time intervals and following the mixed autoregressive moving average model

$$\phi(B)z_t = \theta(B)a_t \qquad (1)$$

where $z_t = Z_t - \eta$, η is a location parameter, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ are polynomials in B have zeros lying on or outside the unit circle, B is the backshift operator such that $Bz_t = z_{t-1}$, and $\{a_t\}$ are a series of white noise, identically and independently distributed as normal with zero means and common variance σ_a^2. When K series

$\{\underset{\sim}{Z}_t\}$, where $\underset{\sim}{Z}'_t = (Z_{1t}, \ldots, Z_{kt})$, are simultaneously considered, a vector analog of (1) is the model

$$\underset{\sim}{\phi}(B)\underset{\sim}{z}_t = \underset{\sim}{\theta}(B)\underset{\sim}{a}_t \tag{2}$$

where $\underset{\sim}{z}_t = \underset{\sim}{Z}_t - \eta$, $\underset{\sim}{\eta}$ is a vector of locations, $\underset{\sim}{\phi}(B) = \underset{\sim}{I} - \underset{\sim}{\phi}_1 B - \ldots - \underset{\sim}{\phi}_p B^p$ and $\underset{\sim}{\theta}(B) = \underset{\sim}{I} - \underset{\sim}{\theta}_1 B - \ldots - \underset{\sim}{\theta}_q B^q$ are matrix polynomials such that the determinantal polynomials $|\underset{\sim}{\phi}(B)|$ and $|\underset{\sim}{\theta}(B)|$ have all zeros lying on or outside the unit circle and $\{\underset{\sim}{a}_t\}$ are identically, independently and normally distributed as normal with zero mean vector and covariance matrix $\underset{\sim}{\Sigma}$.

Summary of Main Results

(i) Analysis and decomposition of seasonal time series:

Economic and environmental time series often exhibit a strong cyclical or seasonal behavior. It has been argued that seasonality should be removed from the series so that the underlying trend becomes more clearly discernable. Suppose that an observable time series Z_t can be written additively as

$$Z_t = S_t + T_t + e_t \tag{3}$$

where S_t is the seasonal component, T_t is the trend component and e_t is the irregular or noise computer. A widely used method is the X-11 seasonal adjustment procedure empirically developed by the U.S. Bureau of the Census over a number of years. It employs specific moving average filters to estimate S_t and T_t from the data Z_t. The main weakness of this

procedure is that the underlying mechanism of S_t and T_t is not clearly specified which makes it difficult to check the appropriateness of the method in any given application. In [1], we have found specific ARMA models of the form in (1) for S_t, T_t and e_t for which the Census filters are approximately optimal. The models for the components then imply an overall model for Z_t which can be verified form the data.

The work in [1] provides a link between an empirically based procedure such as X-11 and procedures which are based on models of Z_t. Model based decomposition and seasonal adjustment procedures are developed in [2], [3], [4] and [5]. In this approach, an overall model for Z_t of the form (1) is first built from the data. Guided by the nature of the model, specific assumptions are then made to decompose Z_t into S_t, T_t and e_t. A canonical decomposition is proposed which minimizes the innovation variances of S_t and T_t. This decomposition is shown to be unique when it exists and has a number of desirable properties. In [5] this procedure is extended to take into account other relevant issues such as trading day and holiday effects as well as removal of spurious observations. Our work has progressed to the point that a large pilot experiment is planned at the Census Bureau to study the performance of the model based procedure with that of X-11, for possible modifications or replacement of the latter in the future.

In modeling seasonal time series, a model of the form (1) has been frequently found useful. It should, however, be noted that models of this kind assume that the time dependent structure among the observations is invariant to shift in origin. This may not be valid in some situations. For example, variability of weather may be different from one season to the next, or the correlation of the demand for a certain item between two

adjacent months may be higher in the summer than in the spring. In [6], properties of a class of periodic models for characterizing such series are explored, and the effects of misspecifying a homogeneous ARMA model (1) for series having a periodic structure are studied.

(ii) Effect of temporal and contemporal aggregation of time series

The problem of temporal and contemporal aggregation of time series data has attracted the interest of many econmetricians. In [7], we have studied the effect on forecasting of temporal aggregation of a univariate series Z_t following the model (1). It is shown that when the length of aggregation is increased, the model for the aggregate approaches a limiting form which depends on the multiplicity of the unit roots of the autoregressive polynomial $\phi(B)$. This result is then exploited to compare the variances of forecasting aggregates using (i) model for the aggregate and (ii) model for the basic series Z_t, and the main conclusion is that there is a substantial gain in using the basic series when it is nonstationary. The implication is that it is worthwhile to use disaggregated data, if available, in forecasting aggregate when the series is nonstationary.

This work is extended in [8] to assess the effect of temporal aggregation on dynamic relationship between two time series variables. We have found that temporal aggregation can turn a unidirectional transfer function relationship into a feedback form. For example, suppose that at the monthly level, consumption linearly depends on income of the last month, and income follows a model of the form (1) independent of past consumption. Thus a unidirectional relationship exists between them. Nevertheless it may be shown that for aggregated quarterly income and consumption data, the

relationship will no longer be unidirectional. Recently, a number of econometricians have attempted to identify causality with unidirectional transfer function relationship found in the data. Our results suggest that extreme care must be exercised in any such attempted interpretation and misleading conclusions can be drawn if the wrong sampling interval is used.

In the problem of contemporal aggregation, we are interested in studying the stochastic structure of a linear aggregate of several series at the same time point, and in comparing forecasts made from past data on the aggregate with those from the component series. Our work in [9] shows the relationship of the stochastic models for the aggregate and the components, and obtains sufficient and necessary conditions for which there will be no gain in accuracy in using the component series to forecast the aggregate.

(iii) Intervention techniques in time series

In analyzing time series data, exogenous interventions often occur, the effect of which is not necessarily immediate on the level of the series. The impact of advertising on the sales of certain product or that of the implementation of automobile exhaust control measures on air quality are examples of such interventions. While much is known about the analysis of data of this kind when the errors are assumed independent and the effects of intervention are immediate, methods developed under such assumptions are not valid for time series data. In [10], it is shown that dynamic effects of intervention may be well represented by difference equation models, and serial dependence of errors may be characterized by (1) (also a difference equation model). The theory and methods are

illustrated by two actual examples, and some further examples are given
in [11].

An alternative analysis of the effects of intervention is given in
[12], in which forecasts made prior to the onset of an intervention are
compared with actual observations since the intervention. This provides
a simple and natural way to formulate models for characterizing the dynamics
of the intervention. Finally, in [13], a detailed study is made on
properties of parameter estimates of a first order dynamic model in
intervention analysis, showing that suitable reparameterization can
appreciably reduce the correlation between parameter estimates in circum-
stances frequently met in practice.

(iv) Analysis and modeling of multiple time series

Economic, engineering and environmental time series data are often
available on several related variables of interest. One may wish to model
and analyze these series jointly in order to (i) understand the dynamic
relationships among these series and (ii) improve accuracy of forecasts.
In the last few years, we have devoted a considerable effort to develop an
iterative approach to modeling multiple time series. An extensive class
of models for such series is the vector ARMA models in (2) which allow
for the possibilities of unidirectional as well as feedback relationships.

In [14], we have extended to the vector case the univariate time series
modeling approach proposed by Box and Jenkins, in Time Series Analysis
Forecasting and Control, Holden-Day, (1970, 1976). A three stage model
building procedure consisting of (a) tentative identification, (b) estimation
and (c) diagnostic checking has been developed and tested on a number of
real examples. In particular, methods for tentative identification employ

cross correlations and partial autoregressions with simplifying indicator symbols to aid in the users' comprehension of the vast amount of information contained in these measures. Parameter estimates are obtained by maximizing the exact likelihood function the properties of which are discussed in [15].

To aid in the understanding of the stochastic structure of multiple time series, various principal component and canonical analyses have been developed. In [16], we have considered linear transformations of the vector Z_t in (2) to characterize the degree of predictability of the data. Transformed series with high degree of predictability represent the growth characteristics in the data, while those with little predictability represent stable relationships. Such analysis has been found useful in uncovering interesting and meaningful substantive structures in the series.

A computer package incorporating the iterative modeling approach as well as the various principal component and canonical analyses has been completed [17]. Over the last 2 years it has been distributed to nearly 200 industrial, academic and governmental institutions throughout the world.

(v) <u>Asymptotic properties of least squares estimates in time series and extended autocorrelation function</u>

For stationary autoregressive model, i.e. in (1) $q = 0$ and all zeros of $\phi(B)$ are lying outside the unit circle, it is well known that the ordinary least squares (OLS) estimates of the autoregressive parameters are consistent and asymptotically normally distributed. However, except for certain special cases, asymptotic properties of the OLS estimates have not been established in the literature for $q \neq 0$ and/or when the series is

not stationary and, largely because of this, no simple method has been available for tentative identification of mixed ARMA models.

In [18], we are able to show the following results. Let $\phi(B) = U(B)\phi(B)$ where $U(B) = 1-U_1B-\ldots-U_dB^d$ has all its zeros <u>on</u> the unit circle and $\phi(B) = 1-\phi_1B-\ldots-\phi_rB^r$ has all its zeros <u>outside</u> the unit circle. Then (a) the OLS estimates of an AR(d) fitting are consistent for the parameters $U_1,\ldots,U_d$, and (b) the OLS estimates of an AR(d+r) fitting are consistent for $\phi_1,\ldots,\phi_p$ in (1) only when $q = 0$.

In [19], an iterated regression method is proposed which yields consistent estimates of $\phi_1,\ldots,\phi_p$ for known q. Since the residuals from the iterated regression approximately follow a moving average model of order q, this has made it possible to define a class of extended auto-correlations as a function of p and q, for tentative specification of the order of a mixed ARMA model. Extension to the vector ARMA model (2) is currently being investigated.

(vi) Pooling of time series

In recent years, there has been a great deal of interest in adopting Bayesian methods for the estimation of means and variance components in random effect models. In [20], such a Bayesian random effect model approach has been applied to the analysis of panel data in a time series context. A detailed analysis is given of a panel of first order auto-regressive models where the autoregressive coefficients across the panel are regarded as a random sample from a beta distribution. Efficiency of pooling the estimates is discussed and illustrated by a real example.

(vii) Effect of outliers in time series

In practice data often contain one or more spurious observations but
these may not be easy to recognize. It is important that statistical
methods are devised which guard against adverse effect of spurious
observations. Methods proposed in the literature for doing this have
frequently made the assumption that the errors are independent which is
unrealistic for time series data. In [21], rules have been proposed to
deal with outliers for a first order autoregressive process. The efficacy
of the rules is studied in relation to estimating the mean and the auto-
regressive coefficient.

(viii) Science and statistics

An important question facing scientists and statisticians is the
precise role that statistics plays in scientific investigation. History
of science shows that scientific progress is achieved largely through a
continued iteration between theory and practice. In [22], a Fisher
Memorial Lecture to the American Statistical Association, the work of
R.A. Fisher is used to illustrate the symbiotic relation of practice and
theory and to show how rapidly progress can occur when the interactive
process is allowed to operate. It is argued that in order to help solve
real problems statistical methods should be developed in this mode.

(ix) Robustness in statistical analysis

An approach to robustification by appropriately modifying the
model and using standard Bayes eatimation is further studied in [25].

In [23] the down weighting patterns applied to "outliers" resulting from such a treatment are compared with those resulting from the Huber Tukey Andrews approach in which the method of estimation rather than the model is modified.

In much of the literature in robust estimation it is tacitly or implicitly implied that errors behave as if randomly drawn from some heavy-tailed distribution. Examination, in [24], of several sets of classical scientific data supposed to support this contention suggest a different situation. It seems that the errors are represented by a roughly normal distribution with a given mean and variance for one period of time and then at a later period of time the variance and/or the mean change to some other values. This will occur for example when pilot experimentation during which the apparatus is being "broken in" is followed by a series of reasonably homogeneous runs during which the apparatus is operating well.

(x) A unified theory of statistical criticism and inference

In [26] it is argued that two different kinds of inference are needed in scientific iteration. Criticism in which the compatibility of model and data is questioned and estimation in which the implications of model and data are combined assuming the compatibility. The view is advanced that ultimately criticism must rest on a sampling theory argument using the predictive distribution as references whereas estimation should be done using the Bayes posterior distribution.

The implications for diagnostic checks, tests of fit and robust and shrinkage estimators is considered.

List of Publications and Technical Reports
(sponsored by the project)

[1] W.C. Cleveland and G.C. Tiao (1976), "Decomposition of seasonal time series: a model for the Census X-11 program", _Journal of the American Statistical Association_, 70, 70-79.

[2] G.C. Tiao and S.C. Hillmer (1978), "Some consideration of decomposition of a time series", _Biometrika_, 65, 497-502.

[3] G.E.P. Box, S.C. Hillmer and G.C. Tiao (1978), "Analysis and modelling of seasonal time series" in _Seasonal Analysis of Economic Time Series_, A Zellner, Ed., 309-333, Bureau of the Census, Department of Commerce.

[4] S.C. Hillmer and G.C. Tiao (1980), "An ARIMA model based approach to seasonal adjustment", Technical Report No. 621, Department of Statistics, University of Wisconsin-Madison, to appear in _Journal of American Statistical Association_, 1982.

[5] S.C. Hillmer, W.R. Bell and G.C. Tiao (1981), "Modeling considerations in the seasonal adjustment of economic time series". Paper for the ASA-Census-NBER Conference on Applied Time Series Analysis of Economic Data, October, 1981.

[6] G.C. Tiao and M.R. Grupe (1980), "Hidden periodic autoregressive-moving average models in time series data", _Biometrika_, 62, 365-373.

[7] G.C. Tiao (1972), "Asymptotic behavior of temporal aggregates of time series", _Biometrika_, 59, 525-531.

[8] G.C. Tiao and W.S. Wei (1976), "Effect of temporal aggregation on the dynamic relationship of two time series variables", _Biometrika_, 63, 513-524.

[9] G.C. Tiao and I. Guttman (1980), "Forecasting contemporal aggregates of multiple time series", _Journal of Econometrics_, 12, 219-30.

[10] G.E.P. Box and G.C. Tiao (1975), "Intervention analysis with applications to economic and environmental problems", _Journal of the American Statistical Association_, 70, 70-79.

[11] G.E.P. Box and G.C. Tiao (1977), "Applications of time series analysis", in _Contribution to Survey Sampling and Applied Statistics_--Papers in Honor of H.O. Hartley, H.A. David, Ed., 203-219, Academic Press.

[12] G.E.P. Box and G.C. Tiao (1976), "Comparisons of forecasts and actuality", _Applied Statistics_, 25, 195-200.

[13] L.M. Liu and G.C. Tiao (1980), "Parameter estimation in dynamic models", _Communication in Statistics_, A9(5), 501-517.

[14] G.C. Tiao and G.E.P. Box (1981), "Modeling multiple time series with applications". _Journal of American Statistical Association_ (to appear in December issue).

[15] S.C. Hillmer and G.C. Tiao (1979), "Likelihood function of stationary multiple autoregressive moving average models", Journal of American Statistical Association, 74, 652-660.

[16] G.E.P. Box and G.C. Tiao (1977), "A canonical analysis of multiple time series", Biometrika, 64, 355-366.

[17] G.C. Tiao, G.E.P. Box, M.R. Grupe, G.B. Hudak, W.R. Bell and I. Chang (1979), The Wisconsin Multiple Time Series (WMTS-1) Program: A Preliminary Guide, Department of Statistics, University of Wisconsin-Madison.

[18] G.C. Tiao and R.S. Tsay (1981), "Consistency properties of least squares estimates of autoregressive parameters in ARMA models", Technical Report No. 658, Department of Statistics, Univ. of Wisconsin-Madison.

[19] G.C. Tiao and R.S. Tsay (1981) "Identification of nonstationary and stationary ARMA models", Technical Report No. 647, Department of Statistics, University of Wisconsin-Madison. (To appear in Proceedings of Annual Meeting of Business and Economic Statistics Section, American Statistical Association, 1981).

[20] G.C. Tiao and L.M. Liu (1080), "Random coefficient first order autoregressive models", Journal of Econometrics, 13, 305-325.

[21] I. Guttman and G.C. Tiao (1978), "Effect of correlation on the estimation of a mean in the presence of spurious observations", Canadian Journal of Statistics, 6, 229-247.

[22] G.E.P. Box (1976), "Science and Statistics", Journal of American Statistical Association, 71, 791-99.

[23] G.G. Chen and G.E.P. Box (1979), "Implied assumptions for some proposed robust estimators", Technical Report No. 568, Department of Statistics, University of Wisconsin-Madison.

[24] ————(1979), "A study of real data", Technical Report No. 569, Department of Statistics, University of Wisconsin-Madison.

[25] ————(1979), "Further study of robustification via a Bayesian approach", Technical Report No. 570, Department of Statistics, University of Wisconsin-Madison.

[26] G.E.P. Box (1979), "Sampling and Bayes' inference in the advancement of learning", Technical Report No. 573, Department of Statistics, University of Wisconsin-Madison.

[27] S.S. Stigler (1973), "The asymptotic distribution of the trimmed mean", Annals of Statistics, Vol. 1, pp. 472-477.

LIST OF PARTICIPATING SCIENTIFIC PERSONNEL

Faculty:

    G. E. P. Box
    J. Ledolter
    S. S. Stigler
    G. C. Tiao

Graduate Students:

Name	Degree/Date
B. Abraham	Ph.D. 8/75
M. Akritas	Ph.D. 8/78
D. Carr	Ph.D. 8/76
R. Chakrabarti	Ph.D. 5/78
I. Chang	(Ph.D. in progress)
G. Chen	Ph.D. 5/79
O. Chiang	M.S. 12/11/74
W. P. Cleveland	Ph.D. 8/72
S. Graves	(Ph.D. in progress)
S. Green	Ph.D. 5/79
M. R. Grupe	Ph.D. 12/79
F. Hernandez	Ph.D. 8/78
S. C. Hillmer	Ph.D. 12/76
H. K. Hsieh	Ph.D. 8/76
J. T. Hsieh	M.S. 4/20/78
G. Hudak	(Ph.D. in progress)
R. Hudlet-Yanez	Ph.D. 5/76
R. Jones	Ph.D. 12/80
H. Kanemasu	Ph.D. 8/73
C. P. Kartha	Ph.D. 8/73
J. Ladalla	Ph.D. 8/76
S. H. Li	Ph.D. 12/75
L. M. Liu	Ph.D. 12/78
S. T. Liu	Ph.D. 12/76
G. Ljung	Ph.D. 12/76
D. Nychka	(Ph.D. in progress)
J. Schaap	M.S. 5/3/73
B. Singh	Ph.D. 12/74
R. S. Tsay	(Ph.D. in progress)
S. C. Wu	Ph.D. 8/79